Coursera | IBM Applied Data Science Capstone

Capstone Final Project

Proposal for a New Vegetarian Restaurant in Hong Kong

Vicky Tsung-En HSU

February 2020

Introduction

Veganism is rapidly growing globally, and Hong Kong is no exception as nearly a quarter of Hong Kong population practised flexitarian (semi-vegetarianism) in 2018[1]. The vegetarian population in Hong Kong grew, from 350k in 2008 to 1.6 million in 2014[2], at a CAGR of 29%. This is contributed by local millennials' values, social media buzz and food scandals.

Hong Kong residents in 2014, in fact, consumed the highest amount of meat and seafood (at 144kg per capita) in the world[3]. However, the consumption habits in Hong Kong have been changing dramatically. The ratio of "hardcore meat lovers" dropped significantly from 27.1% in 2014 to 17.2% in 2016, and further dropped to 15.2% in 2018[1]. Clearly, Hong Kong people are shifting to a plant-based diet.

With the recent trend stated above, the opening of a vegetarian restaurant is a great business opportunity in Hong Kong. But of course, as with any business decision, opening a new venue requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the vegetarian restaurant is one of the most important decisions that will determine whether the restaurant will be a success or a failure.

Business Problem

The objective of the project is to analyse and select the best location in the city of Hong Kong to operate a new vegetarian restaurant. For this purpose, data science methodology and machine learning techniques are executed to figure out a solution. The project targets to answer a very important question: Where would be the best location for a new vegetarian restaurant to be open in Hong Kong?

[1] Green Monday, 2019 (https://greenmonday.org/post/greenmonday-2018-survey-result/)
[2] Statista, 2018 (https://www.statista.com/statistics/975075/hong-kong-number-of-vegetarians/)
[3] Euromonitor, 2015 (https://blog.euromonitor.com/meat-consumption-trends-in-asia-pacific-and-what-they-mean-for-foodservice-strategy/)

Target Audience

Provided that 3.7% of the Hong Kong population are vegetarian[1] suggests Hong Kong, which currently has around 14,000 restaurants[4], should have over 500 vegetarian restaurants to fulfil the demand. Nevertheless, there are less than 300 vegetarian restaurants in Hong Kong[1], implying there is much room for growth. Hong Kong's plant-based grocery and café chain Green Common, for instance, reported the annual sales turnover of 2018 doubled compared with 2017, and has maintained the same growth over the past 3 years[1]. It reflects the continuous growth of demand for plant-based foods in Hong Kong. Also, shopping malls should increase the proportion of vegetarian restaurants to meet the demand of flexitarians.

With the uprising of veganism in Hong Kong, the vegetarian market should be targeted. So, here the project is. As with any retail business, the location is critical for vegetarian restaurants to access more potential customers, develop customer loyalty and especially generate the profits. The target audience then will be very clear – people who share the values or who eager to profit from the trend:

- **Restaurant Owners**: The entrepreneurs who plan to open a new vegetarian business could have a better starting point with an ideal location, while the current restaurant owners could serve the vegetarian options if its location suggests it;
- **Shopping Mall Management** could higher the proportion of vegetarian restaurants if it is located in the recommended area;
- **Restaurant Investors** could make a more comprehensive investment decision with a better understanding of the project;
- **Vegetarian Citizens** or people who value "GREEN" could identify the eco-friendly areas to go, to stay or to live.

[4] FEHD HKSAR (https://www.fehd.gov.hk/english/statistics/pleasant_environment/statistienh_2015_2018.html)

Data

The following data are be required to solve the problem:

- List of Hong Kong's districts. This defines the scope of the project, which is confined to the city of Hong Kong, a Special Administrative Region of the People's Republic of China.

- Latitude and longitude coordinates of those neighbourhoods. The data are used to plot the map and to get the venue data.

- Venue data, particularly data related to vegetarian / vegan restaurants. The data are used to perform clustering on the neighbourhoods.

Sources of data and methods to extract them:

- Wikipedia: "Districts of Hong Kong[5]" includes a list of all districts of Hong Kong, with a total of 18 political areas. The techniques of web scraping are applied to getting the required data from the Wikipedia page, facilitated by the use of Python requests and BeautifulSoup packages. The geographical coordinates of the districts are scrapped from the Wikipedia pages as well.

- Foursquare API: Venue data are extracted from Foursquare location data, which has the dataset of over 62 million venues across over 190 countries with over 900 venue categories[6]. The venue category "Vegetarian / Vegan Restaurant" plays a key role in the execution of data analysis, and further the solution of the business problem.

The project makes use of many data science skills, from web scraping (Wikipedia), interacting with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

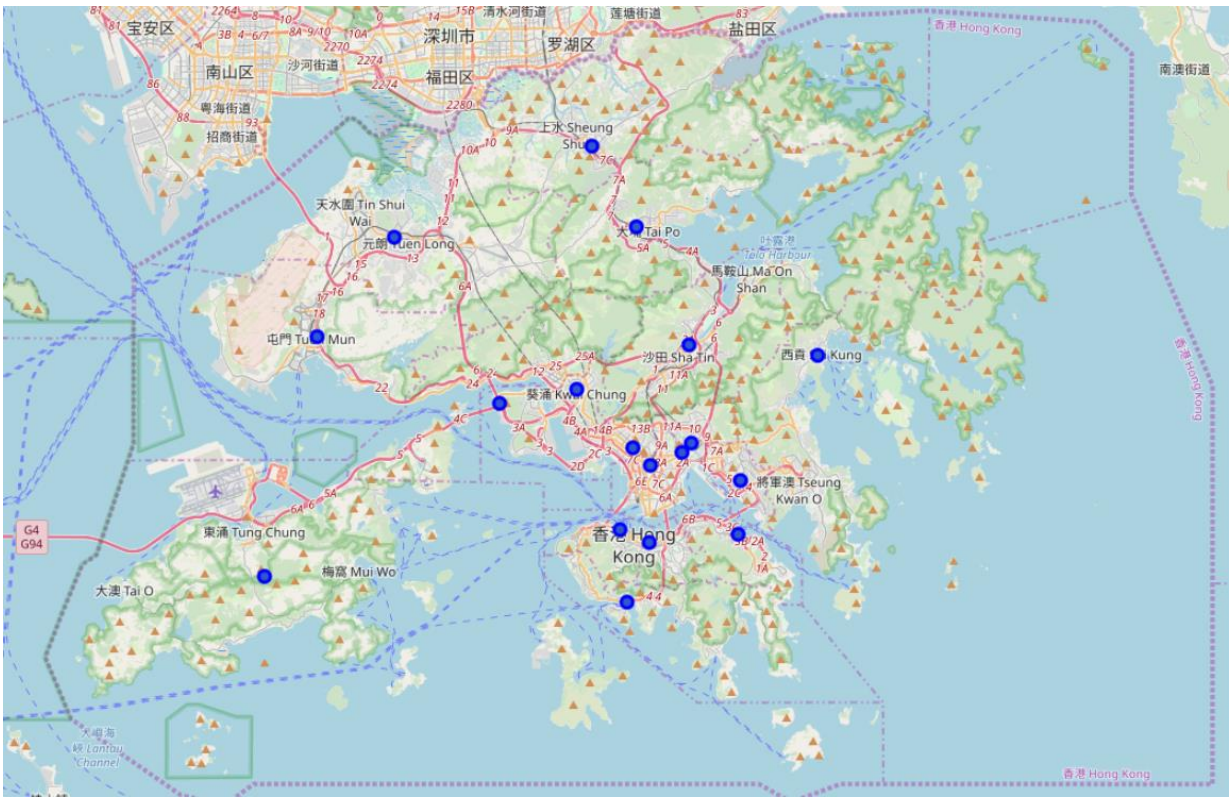
[5] Wikipedia: Districts of Hong Kong (https://en.wikipedia.org/wiki/Districts_of_Hong_Kong)
[6] Foursquare, data extracted on February 2020 (https://enterprise.foursquare.com/products/places)

Methodology

First of all, in order to define the scope of the project, the data of neighbourhoods with the geographical coordinates in Hong Kong are required. The list of districts in Hong Kong and the geographical coordinates are all available on the Wikipedia pages. Yet, the geographical coordinate data are not well-mapped with the list of 18 political areas in the Wikipedia table. The data should be extracted from each individual district page. Hence, applying Python requests and BeautifulSoup packages, I do web scraping to get the list of districts in Hong Kong, and then further scrap the hyperlinks of districts to load the pages and extract the data of latitude and longitude under a Python loop.

After cleaning and sorting the data, the districts with geographical coordinates into a pandas DataFrame is prepared to visualize the districts in a map using Folium package. It can be a sanity check to ensure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Hong Kong. The result is shown as Figure 1.

Figure 1: The map of Hong Kong with blue markers representing the 18 districts



Next, the venue data for the neighbourhoods, obtaining from Foursquare API, are arranged to explore the presence of vegetarian restaurants. At first, the number of venues within a radius of certain meters should be decided for exploring each neighbourhood; hence, a few sets of parameters are tested under a Python loop with Foursquare API calls, until the resulted dataset matches the reality of Hong Kong. After the optimized dataset is generated by appropriate number of venues for each neighbourhood within rational scope, the analysis of the presence of venues can be executed. The density of vegetarian restaurant (the number of "vegetarian / vegan restaurant" venues divided by the number of all "restaurant" venues) by district is used as a key indicator in the project. Thus, the "restaurant" venues are filtered first and the "vegetarian / vegan restaurant" venues are identified then for further data analysis.

Last but not least, K-means clustering is applied to cluster the prepared data. K-means clustering algorithm identifies K number of centroids, and then allocate every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. I cluster the neighbourhoods into three partitions based on the density of vegetarian restaurant. Accordingly, the result allows me to identify the districts with higher or lower concentration of vegetarian restaurant, and it further helps me to answer the defined business question – Where would be the best location for a new vegetarian restaurant to be open in Hong Kong?

Result

The top 100 venues within a radius of 2,500 meters are selected to analyse each neighbourhood, which resulted 1488 venues with 203 unique categories in the city of Hong Kong. Deep diving into the restaurant field, there are 53 unique categories with total 523 restaurants, yet only 11 of them are “vegetarian / vegan restaurants”. Also, as shown in Figure 2, the density of vegetarian restaurant in Hong Kong are still at the low level.

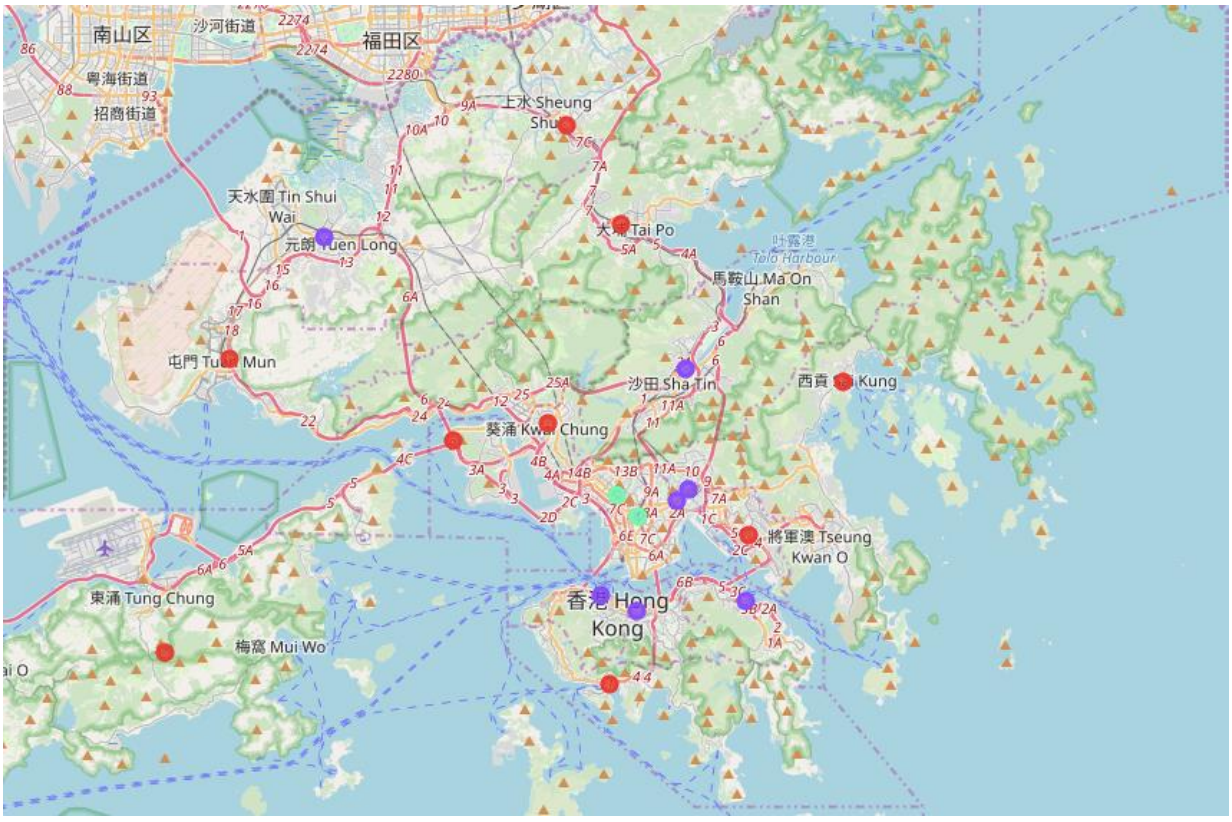
Figure 2: The density of “vegetarian / vegan restaurants” by district in Hong Kong

	Districts	Density of Vegetarian Restaurant
0	Yau Tsim Mong	0.066667
1	Sham Shui Po	0.055556
2	Kowloon City	0.032258
3	Wan Chai	0.032258
4	Wong Tai Sin	0.031250
5	Eastern	0.028571
6	Yuen Long	0.027778
7	Central and Western	0.025641
8	Sha Tin	0.024390
9	Sai Kung	0.000000
10	North	0.000000
11	Southern	0.000000
12	Tai Po	0.000000
13	Tsuen Wan	0.000000
14	Tuen Mun	0.000000
15	Kwun Tong	0.000000
16	Kwai Tsing	0.000000
17	Islands	0.000000

The three clusters are partitioned by the K-means clustering, based on the density of vegetarian restaurant, as Figure 3:

- Cluster 0 (in red colour): Districts with no existence of vegetarian / vegan restaurant;
- Cluster 1 (in green colour): Districts with lower density of vegetarian / vegan restaurant;
- Cluster 2 (in purple colour): Districts with higher density of vegetarian / vegan restaurant.

Figure 3: The clusters by the density of vegetarian restaurant



Vegetarian restaurants are concentrated in the downtown of Hong Kong, while little presence in the suburban regions.

More details please refer to the Notebook on Github:
https://gist.github.com/vickyhte/dd726ee532d0ed1503be7b5654503d74

Discussion

The overall density of vegetarian restaurant is still at the low level in Hong Kong, representing huge potential in Hong Kong green market. Most of the vegetarian restaurants are concentrated in the central area of Hong Kong, with the higher density in cluster 2 (average 6%) and cluster 1 (average 3%), while there is no vegetarian restaurants in the rural area as cluster 0, which includes half of the districts in Hong Kong.

For people who plan to open a new vegetarian restaurant in Hong Kong, cluster 1 is advised to be the best location. First, the vegetarian restaurant density in cluster 2 is double than that in cluster 1, which potentially brings higher competition in the area. Second, there is no vegetarian restaurant in cluster 0 and it may be risky to pioneer an untapped market. Third, considering Hong Kong vegetarians accounting for 3.7% of the population, cluster 1 with the average density of 2.9% still has room to grow, and in fact, all the districts in cluster 1 record the densities under 3.7%. As a result, cluster 1 is an ideal area to be taken into the consideration, especially the districts with lower density in cluster 1. That is, **Sha Tin, Central and Western**, and **Yuen Long**, with the green-food density of 2.4%, 2.6% and 2.8%, respectively.

For the rest of the audience, besides the owners of vegetarian business, here are some suggestions for your reference. **Shopping mall managers** in all areas can consider to increase the proportion of vegetarian restaurants because the green-food market is still at the emerging level for all the districts in Hong Kong. The overall density of 2% is far from satisfying the vegetarian population, 3.7% of the Hong Kong population. Same advice for **the current restaurant owners or investors** – vegetarian cuisine will bring you great opportunities no matter which district you are running your business at. On the other hand, for **vegetarian citizens**, cluster 2, with higher green density, is recommended for you to explore the eco-friendly world. That is, Yau Tsim Mong and Sham Shui Po.

Conclusion

The project is to target on the burgeoning vegetarian market in Hong Kong. The process is gone through defining the business problem, identifying the target audience, specifying the required data, extracting and preparing the data, executing machine learning by clustering into 3 partitions based on the similarities, and lastly providing the advices to the relevant stakeholders regarding the ideal areas to operate vegetarian business in Hong Kong.

To solve the problem and answer the defined question – where would be the best location for a new vegetarian restaurant to be open in Hong Kong? The answer is the districts in cluster 1, especially **Sha Tin, Central and Western**, and **Yuen Long**, which report lower density within the group. The findings of the project can help the relevant stakeholders to make decisions in terms of this high potential market. However, the audience should be noticed that the project only considers one indicator, that is, the density of vegetarian restaurant. There are other crucial factors such as gender ratio, age distribution, education level, average income and consumption behaviour that required the decision markers to examine, in order to figure out an optimal location to operate a new vegetarian restaurant in Hong Kong.